Region-Wise Correspondence Prediction between Manga Line Art Images

Supplementary Material

A. Ablation Study

This section presents the complete quantitative results for the training-scale ablation introduced in Section 5.4 of the main paper. Patch-level and region-level performance on the GenAI dataset is reported in Table A and Table B, respectively. As shown in these tables, scaling up the training data consistently improves structural correspondence, leading to better intra-image region grouping and cross-image region matching.

B. Experiments on 3D-Rendered Line Art

As noted in Section 5.1 of the main paper, in addition to experiments on hand-drawn manga-style line art, we also conduct training and evaluation on the 3D-rendered PaintBucket-Character (PBC) dataset [2]. This section provides details of the dataset conduction, and shows quantitative and qualitative results on the PBC dataset.

B.1. Dataset

The PBC dataset is created by extracting contour lines from 3D-rendered anime characters. This rendering-to-line-art process produces clean and fully closed contours with clearly separable semantic parts, making region segmentation substantially easier than in hand-drawn line art.

The dataset contains over 10,000 synthetic line art images paired with part-level semantic labels under consistent character poses and viewpoints. We sample line art pairs by selecting frames separated by seven frames within the same animation sequence, and define region correspondences based on shared semantic part labels. For example, region `19` in the left image of Figure A and region `40` in the right image share label ID `23`, indicating a semantic correspondence.

In our experiments, we randomly sample 300 image pairs (595 images) for testing, and use the remaining 10,961 pairs (11,336 images) for training.

B.2. Quantitative Results

Patch-level and region-level evaluation results on the PBC dataset are reported in Table C and Table D, respectively.

As shown in Table C, our model achieves near-perfect intra-image patch matching performance, with AP scores above 99.9% and best F1 scores exceeding 98.5% for both I_a and I_b. This demonstrates that the model effectively captures internal region consistency in well-structured synthetic line art.

At the region level, we tune post-processing parameters to preserve visually coherent segmentation while maximizing prediction accuracy. As shown in Table D, our method performs well on both region segmentation and region correspondence, achieving a correspondence accuracy of 59.9%. However, because the PBC dataset contains many small and fine-grained regions, the Cluster Ratio (CR) and region recall remain relatively low, typically around 30%. While reducing over-segmentation and improving region recall represent a fundamental trade-off on this dataset, achieving a better balance will require a deeper structural understanding of abstract line art, highlighting an important direction for future work.

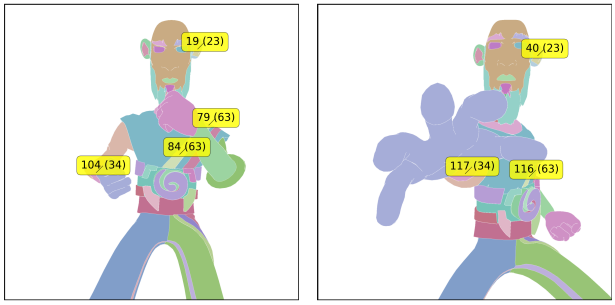

Figure A. Example of region-level correspondence annotations in the PBC dataset. Matched regions are visualized using the same color across two images. Yellow boxes display the region ID and label ID of each region in the format `Region ID (Label ID)`.

B.3. Qualitative Results

Following the qualitative analysis in Section 5.5 of the main paper, we evaluate our method on the PBC dataset from two aspects: (1) ***intra-image region segmentation*** and (2) ***cross-image region matching***. Representative examples are shown in Figure B.

For region segmentation, we compare our method with the Segment Anything Model (SAM) [5]. As observed in Figure B, SAM often merges distinct semantic parts (e.g., face or clothing) with the background and fails to produce fine-grained boundaries due to the lack of color, texture, and shading cues in line art. In contrast, our method produces more precise and semantically coherent region groups.

For region matching, our approach accurately aligns major semantic regions, including cases with noticeable pose variation or shape deformation. The resulting region correspondences generally agree with human perception and demonstrate the applicability of our framework to clean, closed-contour line art.

Training Pairs	Matching Type	AP	Best F1	Precision	Recall	Top-1 Accuracy
20k	Intra-img (I_a)	69.11	64.37	53.23	**81.42**	–
	Intra-img (I_b)	69.23	63.57	53.38	**78.56**	–
	Cross-img	62.17	60.74	48.84	**80.31**	51.58
100k	Intra-img (I_a)	84.82	75.70	**75.90**	75.51	–
	Intra-img (I_b)	85.26	74.34	71.35	77.59	–
	Cross-img	80.72	70.73	68.92	72.65	**68.46**
200k	Intra-img (I_a)	**85.32**	**76.60**	74.69	78.60	–
	Intra-img (I_b)	**86.62**	**76.62**	**77.47**	75.80	–
	Cross-img	**82.53**	**72.79**	**72.96**	72.63	66.86

Table A. Patch-level evaluation on the GenAI dataset under different training set scale.

Training Pairs	Matching Type	ARI	mIoU (P→G)	mIoU (G→P)	CR	Region Accuracy
20k	Intra-img (I_a)	31.32	26.97	35.22	1.94	–
	Intra-img (I_b)	34.83	28.50	35.87	1.90	–
	Cross-img	–	–	–	–	65.72
100k	Intra-img (I_a)	49.53	35.56	**38.58**	1.37	–
	Intra-img (I_b)	54.51	36.93	**40.76**	1.36	–
	Cross-img	–	–	–	–	66.20
200k	Intra-img (I_a)	**50.21**	**38.25**	38.29	**1.26**	–
	Intra-img (I_b)	**55.33**	**39.41**	39.47	**1.20**	–
	Cross-img	–	–	–	–	**70.44**

Table B. Region-level evaluation on the GenAI dataset under different training set scales.

Matching Type	AP	Best F1	Precision	Recall	Top-1 Acc.	Top-5 Acc.
Intra-img (I_a)	99.92	98.57	99.06	98.08	–	–
Intra-img (I_b)	99.93	98.66	98.92	98.40	–	–
Cross-img	98.86	95.23	96.34	94.15	85.44	91.35

Table C. Patch-level evaluation on PBC datasets (percentage values).

Matching Type	ARI	mIoU (P→G)	mIoU (G→P)	CR	Region Accuracy
Intra-img (I_a)	83.05	60.20	20.30	0.32	–
Intra-img (I_b)	83.86	60.34	19.97	0.32	–
Cross-img	–	–	–	–	59.92

Table D. Region-level evaluation on PBC datasets (percentage values; CR is shown in raw value). Region Precision and Region Recall are computed using predicted region pairs with purity > 0.8.

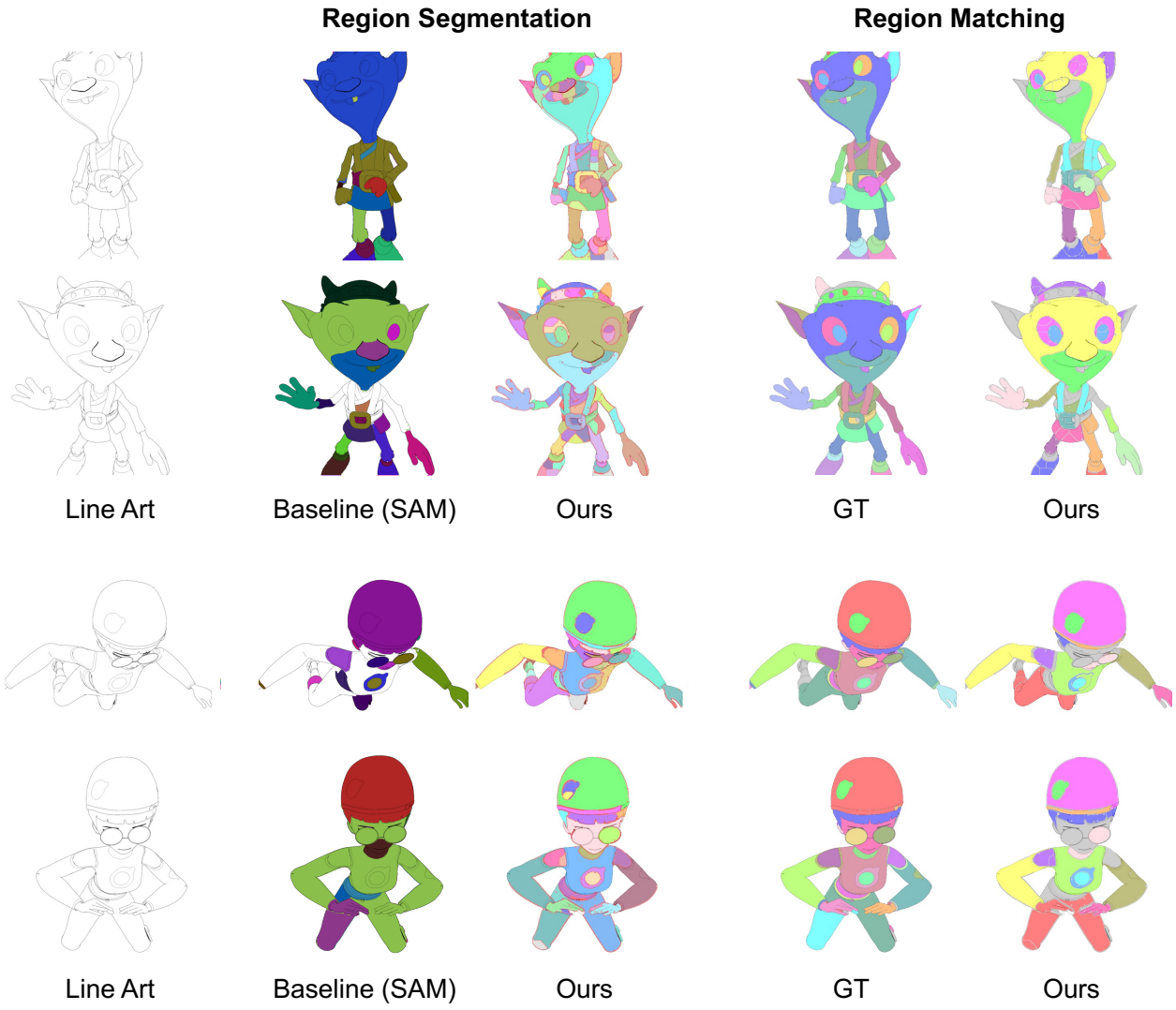

Figure B. Visualization examples on the PBC dataset. In the region segmentation results, the same color within an image denotes a single region, and white indicates the background. In the region matching results, the same color across two images indicates a matched region pair, while gray denotes regions without a match.